1 Prior Predictive Distribution for $\tilde{\mu}^a$

Here, we study the prior predictive distribution of $\tilde{\mu}^a$, the underlying 'true' value of the parameter a for a patient with covariate vector $\tilde{x}$. The goal of this section is to see how our prior belief on what $\tilde{\mu}^a$ varies depends on the hyperparameters. We first claim that in the prior predictive distribution, $\tilde{\mu}^a$ follows a logit-normal distribution, whose properties we will now describe.

1.1 Logit-Normal Distribution

A logit-normal distribution is a continuous distribution defined on the open interval $(0, 1)$. A random variable X follows a logit-normal distribution if the transformed random variable $\text{logit}(X)$ follows a normal distribution. An equivalent definition makes the parameterization of a logit-normal distribution clear:

Definition 1. If $Y \sim \text{Normal}(\mu, \sigma)$, then the transformed random variable $X = \text{logistic}(Y)$ follows a Logit-Normal(μ, σ) distribution.

Thus, a Logit-Normal distributed random variable X is parameterized using the mean and standard deviation of the normal distribution that $\text{Logit}(X)$ follows.

1.1.1 Key Properties

Applying the change of variable formula to the density function of a Normal(μ, σ) random variable under the logistic transformation, one arrives at the density function of a Logit-Normal(μ, σ) random variable:

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(\text{logit}(x) - \mu)^2}{2\sigma^2} \frac{1}{x(1-x)}; x \in (0, 1) \tag{1}$$

Unfortunately, there are no analytical formulas for the mean, variance, or mode of a logit-normal distribution.

1.1.2 Unimodality

In general, the density of a logit-normal distribution may have either 1 or 2 modes. For sufficiently large σ, a Logit-Normal(μ, σ) distribution will have 2 modes - one near 0 and one near 1. This intuitively makes sense, because a sufficiently diffuse normal distribution will have significant mass far away from 0, where the slope of the logistic function is nearly flat. Conversely, for sufficiently small σ, $f_X(x; \mu, \sigma)$ will be unimodal. This result can be seen analytically.

By taking the derivative of $f_X(x; \mu, \sigma)$, it can be seen that the modes of $f_X(x; \mu, \sigma)$ occur at x for which the following condition is satisfied:

$$\text{logit}(x) = \sigma^2(2x - 1) + \mu \tag{2}$$

For any μ and σ, there is always at least 1 x for which this condition is satisfied. Thus, $f_X(x; \mu, \sigma)$ always has at least 1 mode. Furthermore, as the slope of $\text{logit}(x)$ is always greater than 1, we see that if $\sigma^2 < 1$, there is exactly 1 x such that the condition is satisfied, and arrive at the following observation:

Observation 1. If $\sigma^2 \leq 1$, then $f_X(x; \mu, \sigma)$ is unimodal.

1.2 Logit-Normality of $\tilde{\mu}^a$ in the Prior

The only hyperparameter that $\tilde{\mu}^a$ depends on in the prior is $\Sigma^a = c^a I$. Now, we see that $\tilde{\mu}^a|c^a$, the prior predictive distribution of μ^a, follows the logit-normal distribution. This is because $B^a \sim N(0, \Sigma^a)$ and so $B^a\tilde{x} \sim N(0, \tilde{x}'\Sigma^a\tilde{x})$. Then, $\mu_{pop}^{a*} + B^a\tilde{x} \sim N(\mu_{pop}^{a*}, \tilde{x}'\Sigma^a\tilde{x})$. Finally, as $\tilde{\mu^a}|c^a \sim g^a(\mu_{pop}^{a*} + B^a\tilde{x})$ and g^a was defined to be the logistic function, we arrive at the following observation:

Observation 2. $\tilde{\mu}^a|c^a \sim \text{Logit-Normal}(\mu_{pop}^{a*}, \tilde{\sigma}^a)$, where $\tilde{\sigma}^a = \tilde{x}'\Sigma^a\tilde{x} = c^a \sum_{j=1}^{k} \tilde{x}_j^2$

where we have used the fact that Σ^a was parameterized by the hyperparameter c^a.

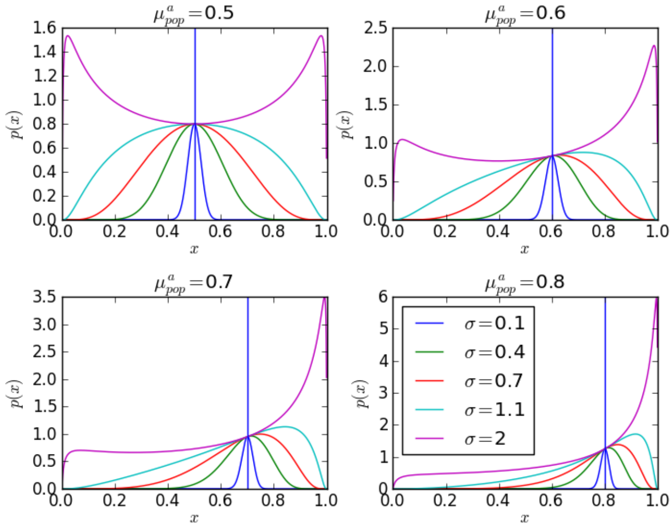

Figure 1: prior predictive distribution of $\tilde{\mu}^a$

1.3 Dependence of Prior Predictive Distribution of $\tilde{\mu}^a$ on Hyperparameters

For a test sample $\tilde{x}$, the prior predictive distribution $\tilde{\mu}^a|c^a$ is distributed Logit-Normal($\mu^{a*}_{pop}, \tilde{\sigma}^a$), where $\mu^{a*}_{pop} = g^a(\mu^a_{pop})$ and $\tilde{\sigma}^a$ is as defined in the previous section. Let us first see how this prior predictive distribution depends on μ^a_{pop} and $\tilde{\sigma}^a$. In Figure X, we have plotted for several values of μ^a_{pop}, how the prior predictive distribution $\tilde{\mu}^a|c^a$ depends on $\tilde{\sigma}^a$. The vertical line denotes the location of μ^a_{pop}.

There are several things to note from the figure:

1. As $\tilde{\sigma}^a$ approaches 0, Logit-Normal($\mu^{a*}_{pop}, \tilde{\sigma}^a$) converges to the point mass at μ^a_{pop}. Recall that we are normalizing the covariate vectors so that each covariate has mean 0 and standard deviation 1 over the training data, so that if a test sample $\tilde{x}$ is equal to the 'average' patient of the training data, then $\tilde{x} = 0$. This means that the more similar a test sample $\tilde{x}$ is to the 'average' patient, the more closer $\tilde{x}$ is to 0, the smaller $\tilde{\sigma}^a$ is, and thus the more strongly we believe in the prior that $\tilde{\mu}^a$ is equal to μ^a_{pop}, the average of a in the training data. This is what we want.

2. If $\tilde{\sigma}^a \leq 1$, then $P(\tilde{\mu}^a; c^a)$ is necessarily unimodal. If $\tilde{\sigma}^a > 1$, then $f(\tilde{\mu}^a; c^a)$ might still be unimodal, but not necessarily.

3. As $\tilde{\sigma}^a$ increases, the mode of $P(\tilde{\mu}^a; c^a)$ increases. While we would like the mode to remain constant as $\tilde{\sigma}^a$ increases, we see this as being unavoidable. Fortunately, the spread of $P(\tilde{\mu}^a; c^a)$ also increases, so that we have a weaker prior belief over $\tilde{\sigma}^a$.

4. The mean of $\tilde{\mu}^a|c^a$ decreases as $\tilde{\sigma}^a$ increases. So the mode and mean exhibit opposite trends.

In Figure X, we plot how the mode of $P(\tilde{\mu}^a; c^a)$ changes with $\tilde{\sigma}^a$, as μ^a_{pop} is held fixed. We do this for several values of μ^a_{pop}.